

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL**

**PREVISÃO DE PRODUTIVIDADE DE SOJA COM MODELOS DE MACHINE
LEARNING**

Maria Gabriela de Sousa

Jaboticabal – SP

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE CIÊNCIAS AGRÁRIAS E VETERINÁRIAS
CÂMPUS DE JABOTICABAL

**PREVISÃO DE PRODUTIVIDADE DE SOJA COM MODELOS DE MACHINE
LEARNING**

MARIA GABRIELA DE SOUSA

Orientador: Prof. Dr. Glauco de Souza Rolim

Trabalho apresentado à Faculdade de Ciências
Agrárias e Veterinárias - UNESP, Câmpus de
Jaboticabal, para graduação em ENGENHARIA
AGRONÔMICA.

Jaboticabal - SP
2º Semestre / 2023

S725p

Sousa, Maria Gabriela

Previsão de produtividade de soja com modelos de machine learning / Maria Gabriela Sousa. -- Jaboticabal, 2023

44 p. : il., tabs., mapas

Trabalho de conclusão de curso (Bacharelado - Engenharia Agrônômica) - Universidade Estadual Paulista (Unesp), Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal

Orientador: Glauco de Souza Rolim

1. Inteligência artificial. 2. Agricultura previsão. 3. Python. 4. Aprendizado do computador. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Ciências Agrárias e Veterinárias, Jaboticabal. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.



UNIVERSIDADE ESTADUAL PAULISTA
CÂMPUS DE JABOTICABAL



DEPARTAMENTO:

Departamento de Engenharia e Ciências Exatas

CERTIFICADO DE APROVAÇÃO

TRABALHO DE INICIAÇÃO CIENTÍFICA

TÍTULO: PREVISÃO DE PRODUTIVIDADE DE SOJA COM MODELOS DE MACHINE LEARNING

ACADÊMICO: Maria Gabriela de Sousa

CURSO: Engenharia Agrônômica

ORIENTADOR (ES): Prof. Dr. Glauco de Souza Rolim

Aprovado e corrigido de acordo com as sugestões da Banca Examinadora

BANCA EXAMINADORA:

(Nomes)

Presidente Glauco de Souza Rolim

Membro Rafael Fausto de Lima

Membro Thifany Stella Cattossi de Miranda

(Assinaturas)

Rafael Fausto de Lima

Thifany Stella Cattossi de Miranda

Jaboticabal 28 / 11 / 2023

Aprovado em reunião do Conselho do Departamento em:

12 / 12 / 2023, por ato "ad referendum".

Prof. Dr. Danisio Prado Munari
Chefe do Departamento

AGRADECIMENTOS

A Deus, o criador de todas as coisas. Pelo dom da vida e pelas realizações durante minha trajetória neste plano.

Aos meus familiares que, mesmo distantes, incentivaram-me e apoiaram-me em todos os momentos, principalmente, aos meus pais Regina Célia D. Calegaro de Sousa e Valdeci de Sousa pelo constante amor e carinho.

Aos amigos, que contribuíram de alguma forma para que eu pudesse chegar até aqui.

Ao meu orientador Glauco de Souza Rolim e coorientador Rafael Fausto de Lima pelos ensinamentos, paciência.

À Universidade Estadual Paulista, especialmente, ao Departamento de Engenharia e Ciências Exatas, pela oportunidade.

À Fundação de Amparo à Pesquisa do Estado de São Paulo – FAPESP, pelo auxílio financeiro para desenvolvimento desta pesquisa - 2023/00166-8.

Aos membros da banca de defesa, pela disponibilidade, conselhos e contribuições ao trabalho.

Muito obrigada!

SUMÁRIO

RESUMO	VII
ABSTRACT	VIII
1 INTRODUÇÃO	09
2 REVISÃO DE LITERATURA.....	11
3 MATERIAL E MÉTODOS.....	18
3.1 Área de estudo e dados	18
3.2 Obtenção de dados meteorológicos.....	20
3.3 Cálculo de evapotranspiração.....	20
3.4 Determinação da influência climática	20
3.5 Modelagem	23
3.6 Avaliação de desempenho dos modelos.....	25
4 RESULTADOS E DISCUSSÃO	25
5 CONCLUSÃO	39
REFERÊNCIAS.....	40

RESUMO

A soja é uma oleaginosa rica em proteínas, principal fonte de óleo vegetal cultivada em diversas regiões do planeta e a sua produção mundial dobrou nos últimos vinte anos. O objetivo deste trabalho é realizar previsão antecipada da produtividade para as principais regiões produtoras de soja do Brasil: Nova Mutum (MT), Sorriso (MT), Formosa do Rio Preto (BA), São Desidério (BA), Itapeva (SP) e Itaberá (SP), a partir de dados meteorológicos e modelos de machine learning. Os dados meteorológicos foram obtidos em escala diária para cada localidade no período de 1989 a 2020. Foram coletados temperatura do ar (máxima, mínima e média), irradiância solar global diária, precipitação, umidade relativa, velocidade do vento, potencial evaporativo, armazenamento de água no solo a partir da plataforma NASA-POWER. A previsão da produtividade foi realizada utilizando os modelos de Regressão Linear Múltipla (RLM), Redes Neurais Artificiais (RNAs), Random Forest (RF) e Extreme Gradient Boosting (XGB), considerando como variáveis independentes as características climatológicas e a variável dependente a produtividade. Para avaliar o desempenho do modelo será usado o coeficiente de determinação ajustado (R^2 ajustado), raiz do erro médio quadrático não-sistemático (RMSEu) e raiz do erro médio quadrático sistemático (RMSEs). Os algoritmos machine learning calibrados apresentaram alta precisão e exatidão tanto na calibração quanto nos testes. O melhor modelo na calibração foi a multi-layer perceptron, que apresentou valores de treinamento para R^2 adjust, RMSEu e RMSEs de 0.99, 0.01 kg/ha e 0.04 kg/ha respectivamente. O melhor modelo no teste foi a MLP, que apresentou valores de R^2 adjust, RMSEu e RMSEs de 0.56, 230.93 kg/ha e 304.763 kg/ha respectivamente.

Palavras-Chave: inteligência artificial, machine learning, previsão de produtividade, python.

ABSTRACT

Soybeans is an oilseed rich in proteins, the main source of vegetable oil cultivated in different regions of the planet and its global production has doubled in the last twenty years. The objective of this work is to carry out an advance forecast of productivity for the main soybean producing regions in Brazil: Nova Mutum (MT), Sorriso (MT), Formosa do Rio Preto (BA), São Desidério (BA), Itapeva (SP) and Itaberá (SP), based on meteorological data and machine learning models. Meteorological data were obtained on a daily scale for each location from 1989 to 2020. Air temperature (maximum, minimum and average), daily global solar irradiance, precipitation, relative humidity, wind speed, evaporative potential, storage of water in the soil from the NASA-POWER platform. Productivity prediction was carried out using Multiple Linear Regression (RLM), Artificial Neural Networks (ANNs), Random Forest (RF) and Extreme Gradient Boosting (XGB) models, considering climatological characteristics as independent variables and productivity as the dependent variable. To evaluate the model's performance, the adjusted coefficient of determination (adjusted R²), unsystematic root mean square error (RMSE_u) and systematic root mean square error (RMSE_s) will be used. The calibrated machine learning algorithms showed high precision and accuracy in both calibration and testing. The best model in the calibration was the multi-layer perceptron, which presented training values for R² adjust, RMSE_u and RMSE_s of 0.99, 0.01 kg/ha and 0.04 kg/ha respectively. The best model in the test was the MLP, which presented R² adjust, RMSE_u and RMSE_s values of 0.56, 230.93 kg/ha and 304.763 kg/ha respectively.

Keywords: artificial intelligence, machine learning, productivity forecasting, python.

1. INTRODUÇÃO

As oleaginosas são consideradas a segunda maior reserva de alimentos em todo o mundo (YUZBASHKANDI; KHALILIAN, 2020). A soja (*Glycine max* (L.) Merrill), originária da China antiga, é uma oleaginosa rica em proteínas e a principal fonte de óleo vegetal cultivada em diversas regiões do planeta. O grão contém cerca de 40% de proteína e 20% de óleo, o que proporciona alta rentabilidade econômica (LIU et al., 2020). A soja pode ser usada como fonte de óleo (refinado para cozinhar e produção de biodiesel), farinha, farelo (alimentação animal), proteína isolada e como alimento in natura (MANDARINO; ROESSING, 2001).

No contexto mundial, a produção de soja dobrou nos últimos 20 anos, passando de 174 milhões de toneladas na safra 2000/2001 para 369,029 milhões de toneladas na safra 2022/2023 com área plantada de 136,029 milhões de hectares (USDA, 2023). Atualmente, um em cada quatro produtos do agronegócio em circulação no mundo é brasileiro e a soja se tornou protagonista (OLIVEIRA et al., 2022)

Na safra de 2022/23 o Brasil foi maior produtor mundial de soja, registrando recordes históricos de área de plantio, produtividade e produção, com 154,566 milhões de toneladas em uma área plantada de 44,062 milhões de hectares, aumento de 6,2% em relação à safra passada. O Estado do Mato Grosso é o maior produtor brasileiro de soja, com 45,600 milhões de toneladas produzidas em uma área plantada de 12,086 milhões de hectares. Já nos EUA, segundo produtor mundial do grão, na safra 2022/23 teve produção de 116,377

milhões de toneladas, área plantada de 34,939 milhões de hectares (EMBRAPA SOJA, 2023; CONAB, 2023).

Para a safra 2023/24, prevê área plantada de soja no Brasil de 45,4 milhões de hectares, e produção em 162 milhões de toneladas, isto representa um aumento de produção de 4,5% em comparação com 2022/23. No MATOPIBA (Maranhão, Tocantins, Piauí e Bahia), as perspectivas são favoráveis para o cultivo da soja (USDA, 2023).

Com preços melhores, o Estado da Bahia prevê aumento na área destinada a produção de soja. O Instituto Mato-Grossense de Economia Aplicada (IMEA) prevê área plantada em 12,22 milhões de hectares na próxima temporada, um aumento de 3% em relação a 2022/23 (USDA, 2023).

Previsões são relevantes e importantes na agricultura, uma vez que quanto mais eficaz a previsão, melhor serão as decisões a serem tomadas com base nos dados (SPANCERSKI; SANTOS, 2021). Prever a produtividade agrícola e os riscos na cadeia produtiva durante os próximos anos, pode auxiliar na quantificação dos impactos provocados pelas mudanças climáticas (TILMAN et al., 2011).

Embora o cultivo de soja tenha importância mundial poucos são os trabalhos que utilizam modelos de *machine learning* na previsão da produtividade, principalmente em função das condições climáticas. Nesse contexto, este trabalho busca o desenvolvimento de modelos agrometeorológicos baseados em ML para previsão antecipada da produtividade da soja, usando as influências climáticas das principais regiões produtoras do Brasil.

2. REVISÃO DE LITERATURA

A soja se adapta melhor em regiões onde as temperaturas oscilam entre 20 °C e 30 °C com ponto ótimo de desenvolvimento em torno de 30 °C. Para a semeadura recomenda-se 25 °C para uma emergência rápida e uniforme. Em regiões com temperatura menores ou iguais a 10 °C são impróprias pois o crescimento vegetativo da soja é pequeno ou nulo (FARIAS; NEPOMUCENO; NEUMAIER, 2007).

Temperaturas acima de 40 °C reduzem a taxa de crescimento, provocam distúrbios na floração, no desenvolvimento de vagens e grãos, acentuando o abortamento de flores e provocando a queda de vagens (VALE; ABDALA, 2018). De maneira geral, a floração da soja é induzida por estímulos termofotoperiódicos. Assim, temperaturas abaixo de 13 °C inibem ou retardam a indução da soja ao florescimento (SEIXAS et al., 2020).

A maturação pode ser acelerada em ocorrência de altas temperaturas e quando associadas com condições de baixa umidade, pode haver o aparecimento de grãos verdes favorecendo danos mecânicos durante a colheita. Temperaturas baixas na maturação com períodos chuvosos ou alta umidade, provocam atrasos na data da colheita pela presença de hastes verdes e retenção foliar (FARIAS; NEPOMUCENO; NEUMAIER, 2007).

A demanda geral de água do cultivo para atingir a produtividade máxima varia de 450 a 800 mm por ciclo, dependendo das condições climáticas, manejo agrícola, duração do ciclo e do estágio de crescimento. Em geral a necessidade

de água na soja aumenta à medida que a planta se desenvolve, atingindo o pico durante os estágios de floração e enchimento dos grãos. (SERAFIM et al., 2012).

A radiação solar além de fornecer energia luminosa para a fotossíntese, fornece também estímulos para diversos mecanismos fisiológicos da soja. A intensidade, duração e características do espectro de luz desempenham um papel integral na determinação das reações morfológicas e fenotípicas exibidas pela soja como arquitetura da planta (altura, engalhamento, número de nós e número de vagens) e o início do processo de floração (SEIXAS et al., 2020).

Como forma de simplificar o entendimento da relação entre as variáveis climáticas e o desenvolvimento da soja, os dados foram agrupados de acordo com os estádios fenológicos do cultivo (Figura 2). A caracterização dos estádios de desenvolvimento da planta de soja proposta por Fehr e Caviness (1977) é a mais utilizada até hoje no mundo inteiro, nesse sistema proposto divide o ciclo de desenvolvimento da planta de soja em dois períodos: vegetativo e reprodutivo.

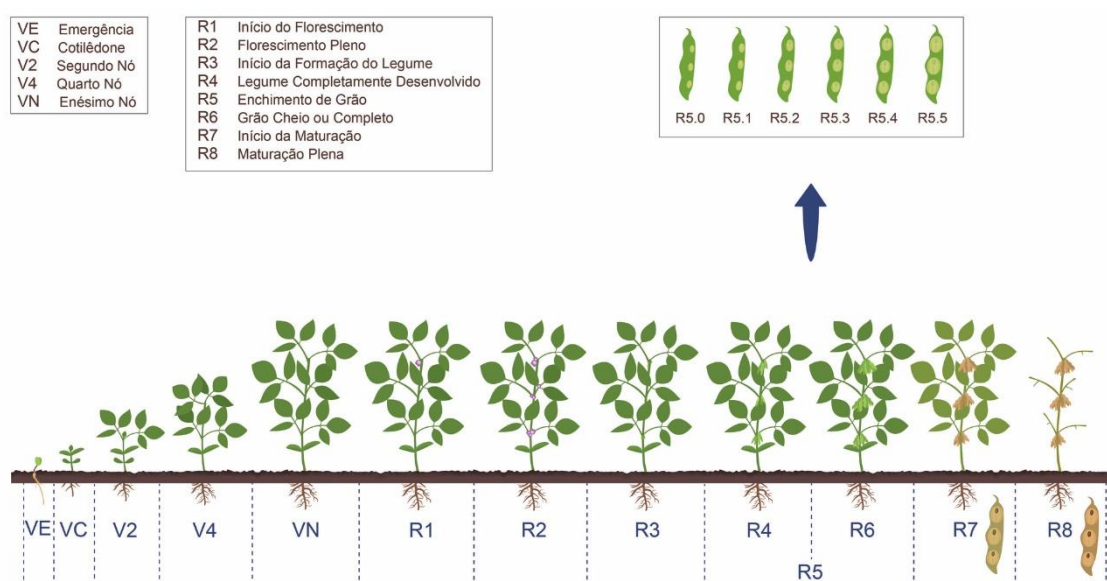


Figura 2 - Estádios fenológicos da soja

Fonte: Tejo, Fernandes e Burrato (2019).

Os estádios vegetativos são designados pela letra V e os reprodutivos pela letra R, com exceção dos estádios VE (emergência) e VC (cotilédone). As letras V e R são seguidas de índices numéricos que identificam estádios específicos, nessas duas fases de desenvolvimento da planta (NEUMAIER et al., 2000). Durante o período vegetativo, a planta cresce em massa e altura, alocando os produtos da fotossíntese às raízes, caules e folhas. As primeiras flores marcam o início do período reprodutivo (R1), quando os produtos da fotossíntese são principalmente alocados aos órgãos reprodutivos (ABRAHÃO; COSTA, 2018).

No estágio VC, a plântula ainda é dependente das reservas dos cotilédones para o suprimento de suas necessidades nutricionais. Até o final do estágio VE, os cotilédones perdem cerca de 70% de seu peso seco. A perda precoce de um dos cotilédones pouco afeta a produtividade final do cultivo, mas

a perda de ambos pode reduzir as produtividades em até 9%. Em V1 a fotossíntese das folhas e a absorção de água e nutrientes pelas raízes da planta em desenvolvimento já são capazes de sustentá-la (RITCHIE et al.,1994).

Em V2 as plantas infectadas pela bactéria *Bradyrhizobium* spp. começam a fixar o nitrogênio do ar. Nos estádios seguintes, a fixação biológica de nitrogênio (FBN) aumenta gradativamente até atingir seu máximo em R5. Em V2, o crescimento das raízes laterais na camada superficial do solo é intenso até o estágio V5. Além disso, perdas drásticas de área foliar durante os estádios vegetativos causam pequenas reduções na produtividade de grãos (QUADROS et al., 2020).

O estágio R3 é marcado pelo início do desenvolvimento das vagens, sendo crucial no estabelecimento do número de vagens por planta, um dos componentes de produtividade mais importantes. Esse componente é sensível às condições de ambiente, provocando abortamento de vagens em condições de estresse (FREITAS, 2019). O estágio R4, também é crucial para determinação da produtividade, ocorrendo um intenso crescimento das vagens e o início do desenvolvimento dos grãos (NEUMAIER et al., 2000).

Consequentemente, qualquer estresse drástico nos estádios de R3 a R5 e início do R6 pode afetar a produtividade de grãos de forma irreversível (TRINDADE, 2018). A disponibilidade de água se torna crucial nessa fase. Metade dos nutrientes necessários para o enchimento de grãos vem da translocação de outras partes da planta, mas a outra metade vem do solo e da

FBN. A necessidade de água pela soja, nesta fase, é de 7 a 8 mm/dia (ALMEIDA, 2022).

No estágio R6, a planta atinge o máximo peso total de vagens, com altas taxas de acúmulo de matéria seca e de nutrientes, tanto na parte aérea (ramos, folhas e caule) como nos grãos. O crescimento das raízes é praticamente nulo ao final do R6 e inicia-se a queda das folhas senescentes dos nós inferiores da haste (FREITAS, 2019).

No R7, os grãos se encontram com cerca de 60% de umidade e já são perfeitamente viáveis como sementes. Estresses, durante e após esse estágio, como chuvas excessivas e ataques severos de percevejos, podem afetar a produtividade e a qualidade dos grãos (ou sementes) (FERREIRA; GOMEZ, 2017). Em R8 maturação plena, há presença de 95% das vagens apresentando coloração de vagens maduras. Mesmo após atingir R8, os grãos ainda necessitam de alguns dias para atingir 13% a 15% de umidade até a colheita dos grãos (NEUMAIER et al., 2000).

A adaptação de diferentes cultivares a determinadas regiões depende, além das exigências hídricas e térmicas, de sua exigência fotoperiódica. A soja é muito sensível ao comprimento do dia, ou melhor, à extensão do período de ausência de luz para a indução floral, sendo considerada planta de dias curtos (TRZECIAK, 2009).

Devido a essa variação, as cultivares são classificadas, quanto à duração do seu ciclo e a latitude, através de grupos de maturidade relativa (GMR).

Portanto, cada cultivar tem uma faixa limitada de adaptação em função do seu grupo de maturidade (ALLIPRANDINI et al., 2009).

A Embrapa Soja apresentou uma proposta de regionalização dos testes de Valor de Cultivo e Uso (VCU) para indicação de cultivares de soja para o Brasil. Os fatores que determinam as macrorregiões sojícolas (MRS), são latitude (fotoperíodo/temperatura) e regime de chuva, enquanto as RECs diferenciam-se por altitude (temperatura) e tipo de solo. Assim foram estabelecidas, MRS e vinte e nove regiões edafoclimáticas (REC) (KASTER; FARIAS, 2012). Em suma, as regiões sojícolas correspondem à diferentes graus de maturações da soja.

A disponibilidade de água, temperatura, radiação solar e fotoperíodo estão entre os principais fatores responsáveis por determinar as máximas produtividades da soja (FARIAS; NEPOMUCENO; NEUMAIER, 2007). Na maioria das vezes, as plantas estão expostas a um elevado grau de estresse, provocado pela ausência de condições ideais de um ou mais fatores ambientais, o que resulta em quebras expressivas de produtividade (SENTELHAS et al., 2017).

Para melhorar os métodos de estimativa de produtividade e produção de soja, há uma necessidade de tecnologias que tornem os sistemas de previsão de safras mais precisos (MONTEIRO; SENTELHAS, 2014).

A utilização do *Machine Learning* (ML), constitui uma das metodologias empregadas no campo da inteligência artificial que usa algoritmos baseados em matemática e estatística para realizar a tarefa de aprendizado da máquina. Esta

tarefa visa extrair informações significativas dos dados, representando-as por meio de um modelo matemático, que pode ser empregado para fazer inferências ou previsões com base em novos conjuntos de dados (MACHADO; MUNDIM, 2023).

Quando adequadamente modeladas, as técnicas de ML fornecem uma melhor capacidade de processamento e análise de dados, facilitando previsões e estimativas (SANTOS, 2020). A previsão difere da estimação na medida em que a estimativa é o cálculo de uma condição atual de várias variáveis atuais, e a previsão é a quantificação de um provável evento futuro (SANTOS; SANTOS; ROLIM, 2021).

Nos últimos anos, os modelos de ML têm sido usados para prever a produtividade dos cultivos, taxa de aplicação de nitrogênio no solo, concentração de nitrogênio nas folhas, classificar sementes, reduzir o fósforo em águas residuais, entre outros estudos (ROLIM, 2000).

O modelo Random Forest, por exemplo, é uma técnica de ML que combina vários algoritmos de árvore de decisão e tem sido usado com sucesso na previsão de safras. Modelos de regressão linear múltipla utilizam várias variáveis independentes, aumentando a chance de previsões de produtividade dos cultivos com maiores acurácias (MENESES, 2020). Outro exemplo são as Redes Neurais Artificiais (RNAs), modelos de aprendizagem profunda que podem ser treinados para analisar e interpretar dados complexos de segurança alimentar, previsões físicas e químicas (SILVA et al., 2022).

Muitos estudos, por exemplo, Peña-Gallardo et al. (2019) usaram resposta da produtividade dos cultivos a diferentes escalas temporais de seca nos Estados Unidos. Arshad et al. (2023) exploraram a resposta dinâmica das secas agrometeorológicas em relação ao risco de perda de produtividade do trigo de inverno usando abordagem de aprendizado de máquina em escala regional no Paquistão. Em Aparecido et al. (2020) propuseram uma modelagem da produtividade do milho safrinha em função das condições climáticas do Mato Grosso do Sul. Miranda et al. (2013) empregaram um modelo matemático para previsão da produtividade do cafeeiro.

3. MATERIAL E MÉTODOS

3.1. Área de estudo e dados

A área de estudo (Tabela 1) abrange os Estados da Bahia (BA) sendo Formosa do Rio Preto e São Desidério, Mato Grosso (MT), como Nova Mutum e Sorriso e também São Paulo (SP) como Itapeva e Itaberá que se destacam sendo as regiões mais importantes em área plantada no Brasil (Figura 1). Foram coletados dados de produtividade obtidos do Sistema de Recuperação Automática do Instituto Brasileiro de Geografia e Estatística (SIDRA-IBGE), no banco de dados da Produção Agropecuária Municipal (PAM), no período de 1989 a 2020, totalizando 31 anos de registros históricos de produtividade de soja.

Tabela 1 - Coordenadas geográficas, altitudes e classificação climática de Köppen dos municípios estudados.

Municípios	Localização Geográfica		Altitude (m)	Köppen
	Latitude (S)	Longitude (W)		
Formosa do Rio Preto (BA)	11° 04' 19"	45° 19' 43"	498	Aw
São Desidério (BA)	12° 36' 08"	44° 97' 03"	492	Aw
Nova Mutum (MT)	13° 84' 44"	56° 08' 16"	452	Aw
Sorriso (MT)	12° 55' 31"	55° 71' 51"	386	Aw
Itapeva (SP)	23° 98' 03"	48° 88' 04"	684	Cfb
Itaberá (SP)	23° 86' 22"	49° 14' 09"	609	Cfb

Fonte: Kaster e Farias (2012); Cidade Brasil (2023).

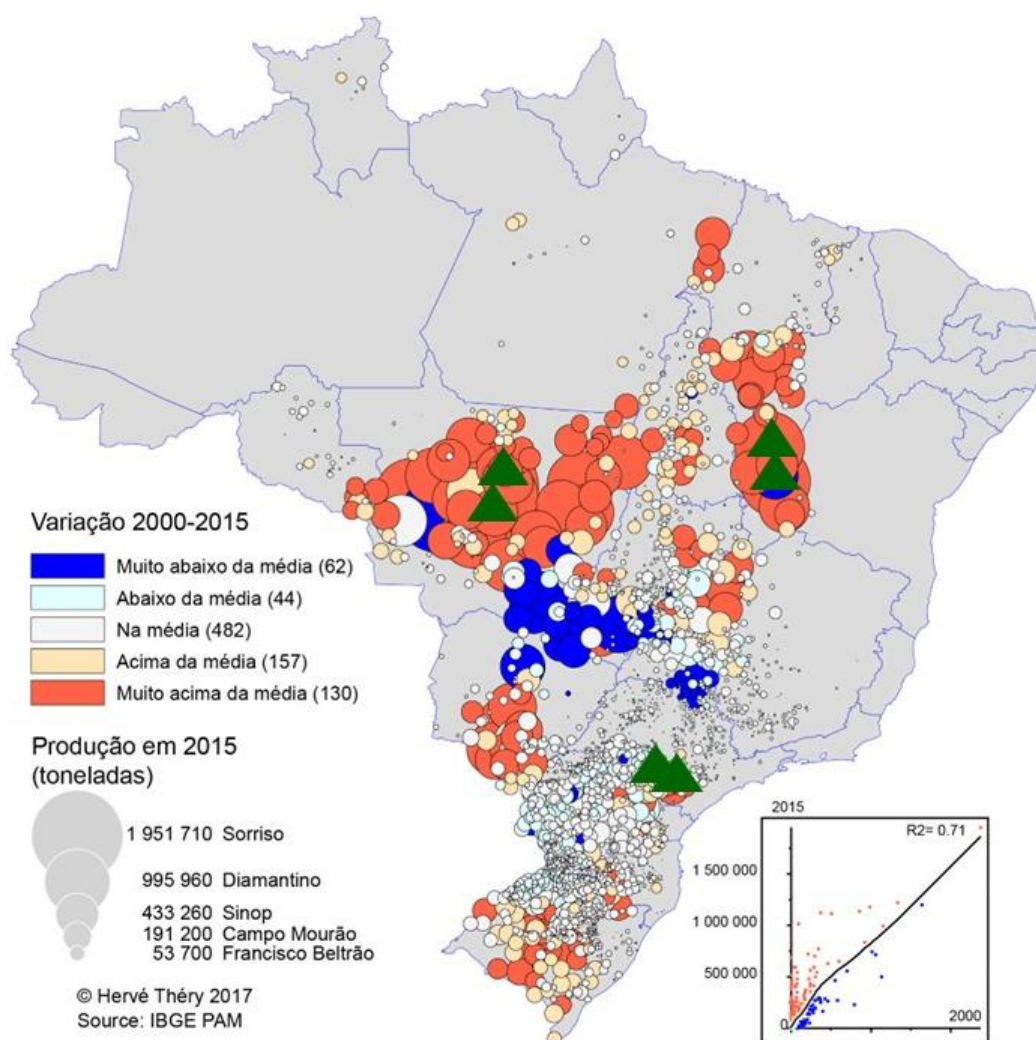


Figura 1 - Áreas produtoras de soja no Brasil compilados por (KNORR, 2017). Simbologia: Δ - locais a serem analisados neste estudo.

3.2. Obtenção dos dados meteorológicos

Os dados meteorológicos foram obtidos em escala diária para cada localidade no período de 1989 a 2020. Os dados coletados foram, temperatura do ar (máxima, mínima e média), irradiância solar global diária, precipitação, umidade relativa, velocidade do vento e armazenamento de água no solo obtidos gratuitamente na plataforma NASA-POWER (<https://power.larc.nasa.gov/>). Os dados diários foram transformados em médias decendiais, excetuando-se a precipitação onde foi realizado o somatório decendial.

3.3. Cálculo de evapotranspiração

A partir dos dados meteorológicos foi calculada a evapotranspiração potencial decendial pelo método de Penman-Monteith (FAO) de acordo com a equação 1 (ALLEN et al., 1998):

$$PET = \frac{0,408\Delta(Rn-G) + \gamma \frac{900}{T+273} U_2 (e_s - e_a)}{\Delta + \gamma(1+0,34U_2)} \quad (1)$$

em que : PET - evapotranspiração potencial (mm dia⁻¹), Rn - saldo de radiação à superfície do cultivo (MJ m⁻² dia⁻¹), G - densidade de fluxo de calor do solo (MJ m⁻² dia⁻¹), T - temperatura do ar em 2 m de altura (°C), U₂ - velocidade do vento a 2 m de altura (m s⁻¹), e_s - pressão de vapor de saturação (kPa), e_a - pressão parcial de vapor (kPa), Δ - declividade da curva de pressão de vapor de saturação (kPa), γ - coeficiente psicrométrico (kPa °C⁻¹).

3.4. Determinação da influência climática

As variáveis agrometeorológicas tem uma grande influência sobre o ciclo de desenvolvimento fenológico do cultivo e com a determinação da produtividade

final. O grau de importância dessas variáveis e a sua influência de em cada estágio do cultivo, foi determinada por meio de análise de correlação. A análise de correlação escolhida foi a de Spearman, uma medida não paramétrica que avalia uma função entre as variáveis com base em sua ordem de classificação, sem a necessidade de normalidade entre os dados (XAVIER et al., 2017).

O desenvolvimento da soja pode ser indeterminado e determinado o que varia de acordo com a adaptabilidade de cada região. No crescimento determinado o período vegetativo cessa com o florescimento, assim o desenvolvimento de vagens e grãos ocorre de forma sincronizada. Cultivares com o tipo de crescimento indeterminado, apresentam na mesma planta crescimento vegetativo, florescimento, formação de vagens e grãos (MARTIN, PIRES; VEY, 2022). No, Brasil, as regiões central e norte utilizam cultivares com tipo de crescimento determinado, mas cada vez é mais crescente o uso de cultivares com tipo de crescimento indeterminado (DECARLI et al., 2022). Estipulamos no trabalho o crescimento determinado como objetivo de estudo.

A metodologia de descrição dos estádios de desenvolvimento proposta por Fehr e Caviness (1977) foram utilizadas para relacionar as interferências das variáveis agrometeorológicas e seus efeitos no ciclo de cultivo. O número de dias considerado em cada estágio fenológico foi: 10 (S – VC); 40 (V1 – Vn); 10 (R1 – R2); 50 (R3 – R6) e 20 (R7 – R8) e 10 (R9).

Apesar da dimensão do território brasileiro e da diversidade edafoclimática, o calendário agrícola para a soja se caracteriza como um cultivo de verão, as regiões da Bahia e Mato Grosso compõem as MRS 4 e REC 405 e

402, respectivamente, e São Paulo MRS 1 e REC 103, com a concentração da semeadura de outubro e novembro. Deste modo, foi escolhido o primeiro decêndio de outubro para o início da semeadura e desenvolvimento da soja nas regiões estudadas, de acordo com o calendário publicado pelo MAPA (Ministério da Agricultura, Pecuária e Abastecimento) e dados disponibilizados em relatórios da Conab (Companhia Nacional de Abastecimento).

As variáveis preditoras (Features) para cada estágio de desenvolvimento da soja, com exceção do estágio R9 (Colheita), foram organizados da seguinte forma: Ts corresponde a temperatura média do ar na emergência, seguindo da mesma forma para Tv12, Tv34, ..., TR8. O mesmo foi aplicado aos demais elementos meteorológicos avaliados (Tabela 2) totalizando 108 features, representando a camada de dados de entradas dos modelos, e a produtividade correspondendo a variável dependente (Target). Nesse estudo, os dados foram divididos em 70% para treinamento e 30% para teste dos modelos.

Características Climatológicas (FEATURES)													
Estádios Fenológicos	Emergência	Vegetativo				Florescimento	Formação de Grãos				Maturação		Colheita
Decêndio	S	V12	V34	V56	V78	R12	R3	R4	R5	R6	R7	R8	R9
Condições climáticas durante o ciclo	OUT-1	OUT-2	OUT-3	NOV-1	NOV-2	NOV-3	DEZ-1	DEZ-2	DEZ-3	JAN-12	JAN-3	FEV-1	FEV-2
	T _s	T _{v12}	T _{v34}	T _{v56}	T _{v78}	T _{r12}	T _{r3}	T _{r4}	T _{r5}	T _{r6}	T _{r7}	T _{r8}	T _{r9}
	T _{maxs}	T _{maxv12}	T _{maxv34}	T _{maxv56}	T _{maxv78}	T _{maxr12}	T _{maxr3}	T _{maxr4}	T _{maxr5}	T _{maxr6}	T _{maxr7}	T _{maxr8}	T _{maxr9}
	T _{mins}	T _{minv12}	T _{minv34}	T _{minv56}	T _{minv78}	T _{minr12}	T _{minr3}	T _{minr4}	T _{minr5}	T _{minr6}	T _{minr7}	T _{minr8}	T _{minr9}
	Q _{gs}	Q _{gv12}	Q _{gv34}	Q _{gv56}	Q _{gv78}	Q _{gr12}	Q _{gr3}	Q _{gr4}	Q _{gr5}	Q _{gr6}	Q _{gr7}	Q _{gr8}	Q _{gr9}
	P _s	P _{v12}	P _{v34}	P _{v56}	P _{v78}	P _{r12}	P _{r3}	P _{r4}	P _{r5}	P _{r6}	P _{r7}	P _{r8}	P _{r9}
	WS _s	WS _{v12}	WS _{v34}	WS _{v56}	WS _{v78}	WS _{r12}	WS _{r3}	WS _{r4}	WS _{r5}	WS _{r6}	WS _{r7}	WS _{r8}	WS _{r9}
	RH _s	RH _{v12}	RH _{v34}	RH _{v56}	RH _{v78}	RH _{r12}	RH _{r3}	RH _{r4}	RH _{r5}	RH _{r6}	RH _{r7}	RH _{r8}	RH _{r9}
	STO _s	STO _{v12}	STO _{v34}	STO _{v56}	STO _{v78}	STO _{r12}	STO _{r3}	STO _{r4}	STO _{r5}	STO _{r6}	STO _{r7}	STO _{r8}	STO _{r9}
	PET _s	PET _{v12}	PET _{v34}	PET _{v56}	PET _{v78}	PET _{r12}	PET _{r3}	PET _{r4}	PET _{r5}	PET _{r6}	PET _{r7}	PET _{r8}	PET _{r9}
Previsão													
													Objetivo (TARGET)
													Produtividade média anual (kg/ha)

Tabela 2 - Esquema de identificação das variáveis de entrada (Features) e saída (Target ou dependente) nos modelos de regressão linear múltipla e random forest a serem testados no projeto. Legenda: T – temperatura média do ar (°C); Tmax – temperatura do ar máxima (°C); Tmin – temperatura do ar mínima (°C); Qg – irradiância solar global diária (MJ m⁻²d⁻¹); P – precipitação (mm); WS – velocidade do vento (m s⁻¹); RH– umidade relativa (%); STO – armazenamento (mm); PET – potencial evaporativo (mm); S – semeadura; V –fase vegetativa; R – fase reprodutiva.

3.5. Modelagem

Para a previsão da produtividade foi utilizado os modelos de Regressão Linear Múltipla (RLM), Redes Neurais Artificiais (RNAs), Random Forest (RF) e o Extreme Gradient Bosting (XGB). A regressão linear múltipla é utilizada em vários trabalhos de modelos preditivos de cultivo que envolve mais de uma variável independente (OBSIE; QU; DRUMMOND, 2020). A RLM é mais realista que a regressão linear simples, devido ao maior numero de variáveis independentes (AIKEN et al., 2012).

Modelos de regressão linear múltipla tem como objetivo obter uma relação matemática entre a variável dependente (variável resposta) e das demais variáveis que descrevem o sistema (variáveis independentes) (FERREIRA, 2016). Os modelos de regressão linear múltipla são mais sensíveis a *outliers*, ocasionando em distorções nos resultados e uma grande diferença nos ajustes (GUJARATI; PORTER, 2011). O maior problema de modelos de regressão linear múltipla é a multicolinearidade, que seria quando duas ou mais variáveis independentes são altamente correlacionadas entre si em um modelo de regressão (FILHO et al., 2011).

A lógica computacional de uma RNA de acordo com Abraham et al., (2017) é semelhante à das redes neurais humanas, eles são treinados a partir de situações que já aconteceram (dados históricos), onde cada neurônio está associado a pesos específicos (intensidade e sinal da conexão) em diversos valores de entrada. As redes neurais são uma técnica de *machine learning*

bastante utilizada na área agrícola, sendo o tipo de redes neurais mais utilizado na previsão da produtividade a Multi-layer Perceptron (MLP) (MENESES, 2020).

A MLP apresenta no mínimo três camadas de nós, que produz um modelo preditivo para uma ou mais variáveis dependentes com base nos valores das variáveis preditoras e é capaz de distinguir dados que não são linearmente separáveis. O erro calculado durante a etapa de treinamento é distribuído pela rede e ajusta os pesos de conexão entre os neurônios (NARAYANAKUMAR; RAJA, 2016).

O RF é composto por um conjunto de árvores que são treinadas com amostras aleatórias do conjunto de dados (FRAJACOMO, 2020). RF é comumente utilizado para classificação e previsão da produtividade dos cultivos (QUIROZ et al., 2018). São combinações de árvores de regressão ou classificação. A ramificação da árvore baseia-se em um subconjunto aleatório de preditores em cada nó e o resultado é determinado como a média de todas as árvores.

O algoritmo RF é executado com eficiência em grandes conjuntos de dados, a precisão é geralmente muito alta, possui método para tratar dados ausentes, fornece estimativa de quais atributos são importantes para regressão e possui uma maior robustez se comparado a outros algoritmos (MARTELETTO, 2022).

XGB é uma das implementações de máquinas de aumento de gradiente que é conhecido como um dos algoritmos de melhor desempenho utilizados para aprendizado supervisionado (OSMAN et al., 2021). Os resultados estatísticos

dos modelos para a previsão da produtividade da soja para os municípios nesse estudo foram divididos em treinamento e teste.

3.6. Avaliação de desempenho dos modelos

O desempenho dos modelos calibrados foi avaliado usando as seguintes estatísticas: o coeficiente de determinação ajustado (R^2 ajustado), raiz do erro médio quadrático não-sistemático (RMSEu) e raiz do erro médio quadrático sistemático (RMSEs) (Eqs. 2-4):

$$R^2 \text{ adjusted} = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1} \right] \quad (2)$$

$$RMSEu = \sqrt{\frac{\sum_{i=1}^N (Y_{est_i} - Y_{est-C})^2}{N}}, Y_{est-C} = a + b \cdot Y_{obs} \quad (3)$$

$$RMSEs = \sqrt{\frac{\sum_{i=1}^N (Y_{obs_i} - Y_{est-C})^2}{N}} \quad (4)$$

na qual, Y_{obs} e Y_{est} são a produtividade observada e estimada, respectivamente.

4. RESULTADOS E DISCUSSÃO

Houve uma grande variabilidade nas condições meteorológicas para o ciclo da soja entre os locais de estudo (Figura 3), favorecido principalmente pela heterogeneidade climática das regiões.

O Estado da Bahia é caracterizado como o mais quente, com Temperatura média de 26,10 ($\pm 1,06$) °C, contra 25,63 ($\pm 0,74$) °C e 21,51 ($\pm 1,05$) °C no Mato Grosso e São Paulo respectivamente (Figura 3A). A mesma tendência pode ser observada para a temperatura máxima (Figura 3B), variando de 28,48 °C a 34,65 °C na Bahia. Com relação as mínimas (Figura 3C) o estado

de São Paulo apresentou os menores valores com média de $16,78 (\pm 0,64) ^\circ\text{C}$ durante o ciclo de desenvolvimento da soja. Os municípios estudados em SP estão dispostos em altitudes maiores e latitudes menores, proporcionando temperaturas mais amenas, em comparação com os demais locais (NOVAIS, 2023).

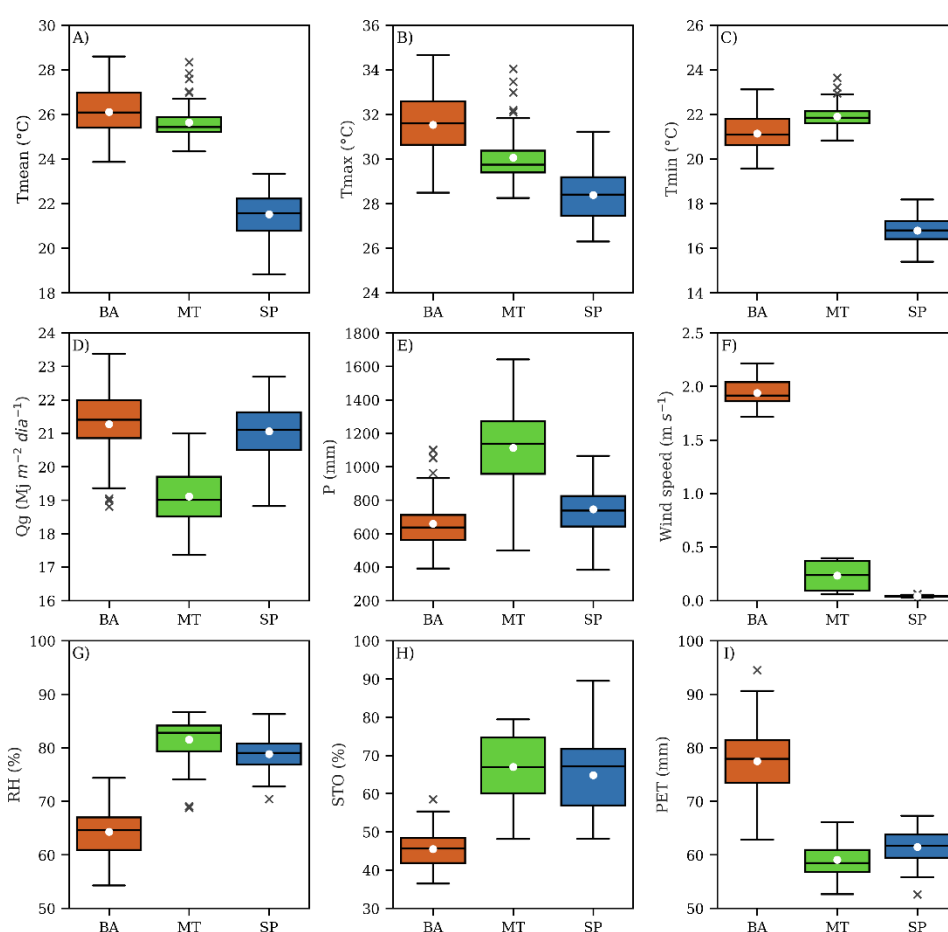


Figura 3 - Médias das variáveis climáticas para o ciclo de desenvolvimento da soja. Legenda: Os eixos Y correspondem as Médias de : A) Temperatura média do ar ($^\circ\text{C}$), B) Temperatura máxima do ar ($^\circ\text{C}$), C) Temperatura mínima do ar ($^\circ\text{C}$), D) Irradiância solar global diária ($\text{MJ m}^{-2} \text{d}^{-1}$), E) Precipitação (mm), F) Velocidade do vento (m s^{-1}), G) Umidade relativa (%), H) Armazenamento (%), I) Potencial evaporativo (mm) e o eixo X corresponde às regiões. (— interno) representa o segundo quartil ou mediana; (□) primeiro e terceiro quartis; (T and ⊥) terceiro e primeiro quartis até os limites superior e inferior; (x) outliers.

Para a radiação solar, o MT apresentou a menor quantidade de energia incidente na superfície (Figura 3D) média de $19,11 (\pm 0,76) \text{ MJ m}^{-2} \text{ dia}^{-1}$ em todo o ciclo do cultivo. Valores superiores foram obtidos em SP e BA com $21,05 (\pm 0,87) \text{ MJ m}^{-2} \text{ dia}^{-1}$ e $21,27 (\pm 0,96) \text{ MJ m}^{-2} \text{ dia}^{-1}$ respectivamente. A radiação solar influencia na fotossíntese sendo assim fundamental para o acúmulo de biomassa e desenvolvimento produtivo da soja (MONTEIRO, 2009).

Em relação a precipitação (Figura 3E), os maiores volumes foram registrados no MT com acumulado médio de $1112 (\pm 244) \text{ mm}$. Vários estudos confirmam um período chuvoso no verão, na qual uma massa equatorial continental age sobre a região (Borges et al., 2023; Arvor et al., 2007). Em SP e BA os acumulados de precipitação foram semelhantes com médias de $745 (\pm 145) \text{ mm}$ e $659 (\pm 155) \text{ mm}$ respectivamente. Esses acumulados estão dentro da faixa adequada ao desenvolvimento da soja, na qual exige 450 a 800 mm durante o ciclo (SERAFIM et al., 2012).

A velocidade do vento (Figura 3F) obteve variação de $0,02 \text{ m s}^{-1}$ a $2,21 \text{ m s}^{-1}$ entre as localidades estudadas, sendo mais elevada na BA com média de $1,93 (\pm 0,11) \text{ m s}^{-1}$. Para a umidade relativa (Figura 3G) a menor média foi obtida na BA com $64,3 (\pm 4,85) \%$. MT e SP registraram umidade relativa média de $81,5 (\pm 4,05) \%$ e $78,8 (\pm 3,12) \%$ respectivamente.

Longos períodos com umidade acima de 75% como registrados em SP e MT favorecem o aparecimento de doenças, que podem resultar em perdas de 15% a 20% na produtividade, porém, se tratando da Ferrugem Asiática da Soja (FAS), principal doença da soja no Brasil (OLIVEIRA et al., 2022; AKAMATSU et

al., 2013; LANGENBACH et al., 2016;), os danos podem acarretar perdas de até 90% de produtividade (ESCOBAR et al., 2023).

A umidade do solo (STO) e a evapotranspiração potencial (PET) apresentaram características inversamente proporcionais (Figura 3H I). A Bahia registrou o menor percentual médio STO e os maiores acumulados de PET no ciclo da soja com 45,5 ($\pm 4,7$) % e 77,4 ($\pm 6,5$) mm respectivamente. Em contrapartida, no MT foram registrados maiores STO e menores PET com médias de 66,9 ($\pm 8,21$) % e 59,07 ($\pm 3,01$) mm respectivamente. Maiores valores de STO estão relacionados aos maiores acumulados de precipitação e a PET elevada com as maiores TMEAN

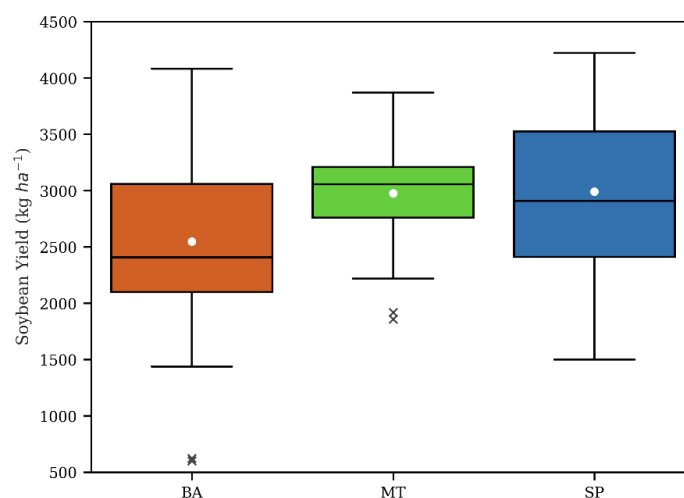


Figura 4 - Produtividade das regiões estudadas. Legenda: O eixo Y corresponde a produtividade e o eixo X corresponde as regiões. Legenda: (— interno) representa o segundo quartil ou mediana; (□) primeiro e terceiro quartis; (T and ⊥) terceiro e primeiro quartis até os limites superior e inferior; (x) outliers.

Em relação a produtividade (Figura 4), as localidades em estudo apresentaram médias próxima entre os três estados, com 2549 (± 751) kg/ha na BA, 2976 (± 396) kg/ha no MT e 2993 (± 717) kg/ha em SP. A baixa produtividade

média dessas regiões ocorre devido ao longo período de dados utilizados 1989 – 2020, isso explica a variação de 1500 kg/ha a 4000 kg/ha na BA e em SP. Esse incremento deve-se ao desenvolvimento de variedades de soja mais produtivas, resistentes a pragas e doenças e adaptadas a diferentes condições climáticas, e também na adoção de práticas agrícolas avançadas e tecnológicas (FERREIRA, 2020).

Na safra 2022/23 a produtividade média da soja no Brasil foi de 3.508 kg ha⁻¹ (58 sc ha⁻¹) (CONAB, 2023). As melhores produções registradas ficam para as regiões de São Paulo com a máxima de 4.223 kg ha⁻¹ no ano de 2016 em Itapeva. Bahia com 4.084 kg ha⁻¹ em 2020 em São Desidério. Em Sorriso, Mato Grosso a produtividade foi de 3.870 kg ha⁻¹.

A correlação entre as variáveis climáticas para cada fase fenológica (Figura 5) e a produtividade da soja para os estados apresentam características distintas. As regiões da Bahia (Formosa do Rio Preto e São Desiderio) são mais impactadas pelo clima, com maior evidencia a partir do estágio fenológico R4, seguido pelas regiões do Mato Grosso e São Paulo.

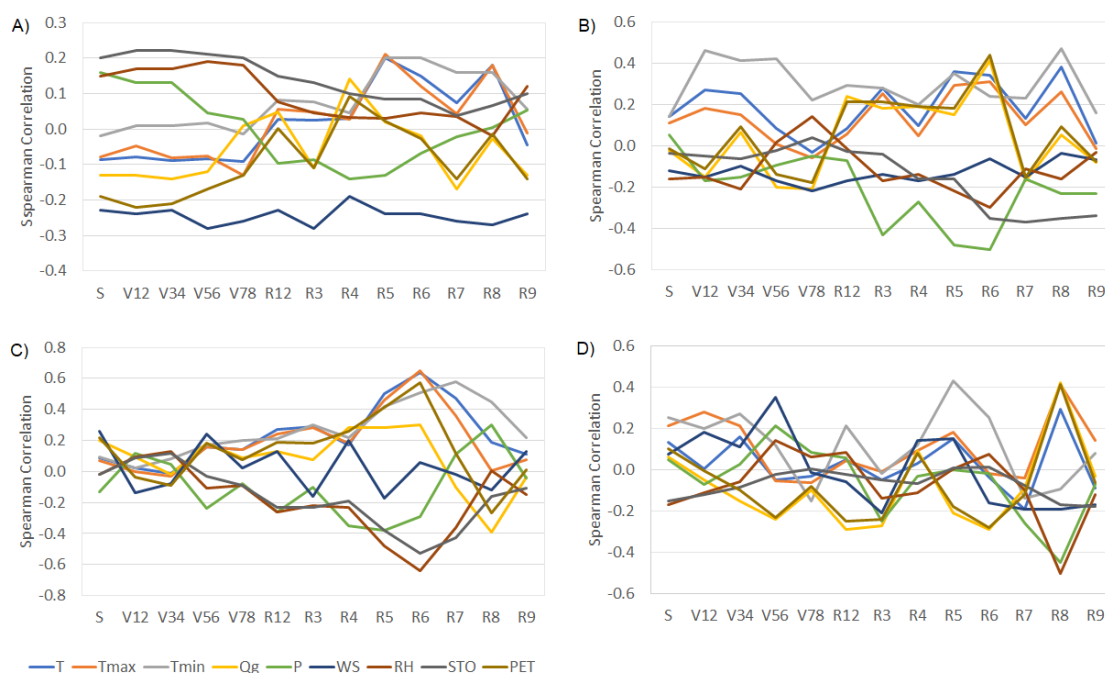


Figura 5 – Correlações (Spearman) da produtividade com variáveis meteorológicas em cada estágio fenológico em: A) todas as regiões (MT, BA e SP), B) Mato Grosso, C) Bahia e D) São Paulo. Legenda: T – temperatura média do ar ($^{\circ}\text{C}$); Tmax – temperatura do ar máxima ($^{\circ}\text{C}$); Tmin – temperatura do ar mínima ($^{\circ}\text{C}$); Qg – irradiância solar global diária ($\text{MJ m}^{-2}\text{d}^{-1}$); P – precipitação (mm); WS – velocidade do vento (m s^{-1}); RH – umidade relativa (%); STO – armazenamento (mm); PET – potencial evaporativo (mm); S – semeadura; V – fase vegetativa; R – fase reprodutiva.

Para todas as localidades estudadas (Figura 5A) a correlação entre as condições meteorológicas de temperatura média e temperatura máxima foi negativa durante a semeadura e o período vegetativo (-0,1). Essas regiões já apresentam uma faixa ótima de temperaturas entre S a R12 com média de 25,3 $^{\circ}\text{C}$ e máxima 31,3 $^{\circ}\text{C}$, indicando que um aumento afetaria diretamente na germinação, emergência e também no crescimento e desenvolvimento da soja.

A correlação foi positiva (0,3) durante o estágio reprodutivo pois a partir do terceiro decêndio de novembro as temperaturas ficam mais amenas nas regiões com média de 24 $^{\circ}\text{C}$ e máxima de 29,1 $^{\circ}\text{C}$. Nesse período também fica mais chuvoso conforme representado no gráfico com correlação negativa (-0,1).

Na região do Mato Grosso as temperaturas têm forte influência sobre todo ciclo da soja, devido as correlações positivas. Destacando a T_{min} com correlação de 0,2 durante a fase vegetativa. T_{min} mais elevadas podem trazer efeitos benéficos na produtividade final da soja. Nesse período temperaturas baixas podem atrasar a emissão de novas folhas, favorecendo, assim também a competição com plantas daninhas e prejudicando o aproveitamento de água e nutrientes (FOERSTER, 2021).

Durante o período reprodutivo, a P e o STO demonstraram correlação negativa entre -0,4 e -0,5, indicando que um aumento no volume de chuvas promoverá efeitos negativos na produtividade, uma vez que o armazenamento de água no solo é elevado nesse período de dezembro a fevereiro, o mais chuvoso na região.

Para a localidade da Bahia as temperaturas mais amenas a partir do estágio R5 proporcionaram correlação positiva entre 0,2 e 0,6. Devido aos altos volumes de chuva nos estádios R4 a R6, a correlação é negativa durante esse período, favorecendo o armazenamento de água no solo e a umidade relativa do ar.

Para Itapeva e Itaberá, os picos de correlação se tornam aparentes em R8 durante os dez dias iniciais de fevereiro, período em que há uma correlação positiva para a temperatura média, temperatura máxima e irradiância. Esse período corresponde ao o estágio final de maturação, e uma escalada nessas variáveis climáticas seria benéfica para atingir a umidade de colheita desejada. Nesse momento, a planta já concluiu o transporte de nutrientes para os grãos e

atingiu o ápice da matéria seca, tendo uma perda gradual de água (MONTEIRO, 2009).

A correlação do potencial evaporativo também é positiva, nesse momento um PET alto teria mais efeitos benéficos para a soja. A correlação de precipitação mostrou-se negativa no estágio R8, indicando que um aumento no volume de chuva promove efeitos negativos na produtividade, dado que os índices desse período são suficientes para suprir as necessidades do cultivo.

É importante enfatizar que o conhecimento das condições meteorológicas de cada região e o estágio são fundamentais para o sucesso do cultivo da soja. Na semeadura a disponibilidade de água é fundamental para o crescimento inicial da soja, no entanto, apenas nas regiões da Bahia não houve condições de precipitações favoráveis, porém todas as localidades apresentaram correlação negativa de armazenamento de água no solo, indicando um cenário favorável para o estabelecimento inicial do cultivo.

Para minimizar os efeitos do déficit hídrico, indicam-se a adoção de práticas culturais como a semeadura em época recomendada que coincida com período de menor risco climático, seguindo indicações do zoneamento agroclimático. Além disso, deve-se adotar práticas que favoreçam o armazenamento de água pelo solo, como o controle de invasoras e a adoção de sistemas de cultivo que resultem na manutenção ou aumento da matéria orgânica e elevada cobertura do solo com palha (FARIAS; NEPOMUCENO; NEUMAIER, 2007).

Já no período vegetativo da soja que compreende a maior parte do ciclo de crescimento da planta, vários fatores climáticos desempenham um papel fundamental no desenvolvimento do cultivo que são eles: temperatura do ar, radiação solar e chuva.

A temperatura juntamente com a irradiância solar e o fotoperíodo afeta uma série de processos na planta como a respiração de manutenção, transpiração, duração das fases fenológicas, indução ao florescimento, crescimento da biomassa, o conteúdo de óleo nos grãos, etc (PEREIRA; ANGELOCCI; SENTELHAS, 2002; SENTELHAS et al., 2017). A temperatura influencia diretamente na fotossíntese, afetando a taxa das reações metabólicas das plantas, regulando o crescimento e o desenvolvimento vegetal. Temperaturas e a irradiância crescentes induzem ao aumento da taxa de fotossíntese, invertendo-se, sob temperaturas muito elevadas (MONTEIRO, 2009). De forma geral, as temperaturas mantiveram correlação positiva durante V12 a V78 com médias na faixa ideal entre 20 °C e 30 °C.

Já as precipitações em períodos de secas induzem as plantas ao fechamento de seus estômatos, fixando menos CO₂, afetando negativamente a fotossíntese. Por outro lado, períodos com chuvas excessivas levam à redução da oxigenação dos solos, diminuindo a atividade radicular e a absorção de água e nutrientes pelas plantas. Tanto as secas quanto o encharcamento dos solos levam à redução da produtividade dos cultivos (MONTEIRO, 2009). As localidades mais impactadas pelas perdas na produtividade devido à falta de chuva no estágio vegetativo são Itapeva e Itaberá, seguido de Formosa do Rio

Preto e São Desiderio. Situação contrária ao observado nas regiões do Mato Grosso onde as precipitações são suficientes ou até excessivas.

O florescimento é um estágio crítico no ciclo de crescimento da soja, pois afeta diretamente a formação de vagens e, conseqüentemente, a produtividade do cultivo. A soja é uma planta sensível ao fotoperíodo, demonstrando uma relação direta com a quantidade de luz que a planta recebe, sendo crucial para estimular a produção de flores e a polinização. Sob radiação solar máxima, a soja atinge máximas produtividades desde que bem suprida de água (THOMAS; COSTA, 2010).

Com relação à temperatura, existe uma relação inversa entre a temperatura média e o número de dias necessários para a floração. Dessa forma, temperaturas mais baixas limitam o crescimento e o desenvolvimento das plantas aumentando o período para que ocorra o florescimento. Além disso, as altas temperaturas promovem danos à floração, reduzindo a capacidade de retenção de legumes, sendo intensificado pelo déficit hídrico (RODRIGUES et al., 2000; SENTELHAS et al., 2017).

Essas condições para o florescimento (R12), no terceiro decêndio de novembro, foram mais bem observadas nas regiões de São Paulo, em que a irradiância é mais alta e as temperaturas são mais amenas com precipitações e armazenagem de água no solo suficientes.

No período reprodutivo (R3-R6) a disponibilidade de água é fundamental, sendo o mais sensível ao déficit hídrico. A necessidade de água no cultivo

aumenta conforme o seu desenvolvimento, atingindo o máximo no florescimento e enchimento de grãos e decrescendo posteriormente (BARBOSA et al., 2015).

A soja quando submetida a períodos de déficit hídrico durante o enchimento de grãos, os grãos são reduzidos em tamanhos e peso e há a retenção da cor verde, senescência foliar e encurtamento do ciclo de vida da planta, devido a diminuição do suprimento dos fotoassimilados ou a inibição do metabolismo do próprio grão (BORRMANN, 2009; ROSA et al., 2020). Todas as regiões apresentaram correlação negativa ou próximo de zero de P e STO indicando condições apropriadas de disponibilidade de água nessa fase.

Ao contrário da formação de vagens, o enchimento de grão no período da maturação a colheita da soja precisa de condições mais secas. Sendo os fatores mais importantes a temperatura, umidade e precipitação, que influenciam diretamente o processo de perda e ganho de umidade dos grãos, o que resulta em lesões do tegumento da semente e em consequente decréscimo da produtividade e da qualidade fisiológica dos grãos (TSUKAHARA et al., 2016). Os resultados das correlações nesse período apresentaram características distintas para cada região. Em São Paulo e Mato Grosso um aumento da temperatura em R8 seria favorável, e na Bahia um aumento na temperatura e na precipitação favorece por se tratar de uma localidade com poucas chuvas.

A modelagem para previsão da produtividade da soja apresentou resultados distintos entre a calibração e teste dos modelos de ML e a regressão linear múltipla em termos de precisão e acurácia. A tabela 3 contém os parâmetros testados e definidos de cada algoritmo. Os modelos ajustados

proporcionaram realizara a previsão da produtividade da soja com antecedência mínima de 10 dias até a colheita.

Tabela 3 - Parâmetros dos modelos de machine learning utilizados na biblioteca scikit-learn. Símbolos: RLM (Regressão Linear Múltipla), MLP (Multi-layer Perceptron), RF (Random Forest) e XGB (XGBoost).

Modelo	Parâmetro	Valores testados	Valores definidos
RLM	(seleção de variáveis por stepwise)		
MLP	hidden_layer_sizes	(5, 10); (40); (10, 3); (15, 25); (25, 3); (25, 75); (80); (90;60)	80
	activation	tanh; relu	relu
	solver	sgd; adam; lbfgs	lbfgs
	alpha	0.01; 0.1; 0.05; 0.5; 0.7; 0.0001;	0.7
	random_state	9	9
	max_iter	500; 1000; 10000; 15000	15000
RF	random_state	9	9
	max_features	0.5; 0.9; 1.0;	0.9
	min_samples_leaf	1; 2; 5; 6; 7; 8; 10; 15; 20	1
	min_samples_split	2; 5; 7; 15; 20;	2
	n_estimators	150; 200; 250; 500; 1000; 1500	1500
XGB	objective	reg:squarederror	reg:squarederror
	learning_rate	0.01; 0.03; 0.05; 0.10; 0.15	0.15
	colsample_bytree	0.1; 0.3; 0.4; 0.5; 0.7; 0.9	0.9
	colsample_bylevel	0.1; 0.3; 0.4; 0.5; 0.7; 0.9	0.9
	colsample_bynode	0.1; 0.3; 0.5; 0.7; 0.9; 1	1
	max_depth	0; 1; 2; 3; 4; 5; 6; 8; 10	0
	min_child_weight	1; 3; 5; 7; 10; 15	15
	subsample	0.3, 0.5; 0.8	0.5
	n_estimators	500; 1000; 2000; 10000	2000

Os melhores modelos na calibração foram a MLP que apresentou valores de treinamento para R^2_{adjust} , RMSEu e RMSEs de 0.99, 0.01 kg ha⁻¹ e 0.04 kg ha⁻¹ respectivamente, e o XGB com 0.99, 0.01 kg ha⁻¹ e 0.02 kg ha⁻¹ para o

R^2 adjust, RMSEu e RMSEs respectivamente (Tabela 4). O XGBoost implementa árvores de decisão com gradiente ampliado que são eficientes e altamente otimizadas para prever a produtividade das safras.

O Random Forest foi o algoritmo com menor precisão no treinamento, apresentando valores de R^2 adjust, RMSEu e RMSEs de 0.49, 191.65 kg ha⁻¹ e 78.08 kg ha⁻¹ respectivamente (Tabela 4). A RLM é um método simples e por isso muito utilizado para previsões. Porém, quando o conjunto de dados de variáveis independentes apresenta maior quantidade de amostras ou possui multicolinearidade, o método não é bem-sucedido (JUNIOR, et al., 2016).

De forma geral, os modelos de aprendizado de máquina usados para prever a produtividade da soja demonstraram bons resultados em testes com dados independentes. O melhor modelo no teste foi a MLP, que apresentou valores de R^2 adjust, RMSEu e RMSEs de 0.56, 230.93 kg ha⁻¹ e 304.763 kg ha⁻¹ respectivamente (Tabela 4).

A RLM foi o algoritmo com menor precisão, com valores de R^2 adjust, RMSEu e RMSEs de 0.22, 125.73 kg ha⁻¹ e 370.469 kg ha⁻¹ respectivamente. A relação de desempenho de cada algoritmo é mostrada na Figura 6. Esse baixo desempenho da RLM também foi observado por Fishman et al. 2010.

Tabela 4 - Estatísticas de avaliação do desempenho na previsão de produtividade.

Modelo	Cross Validation					
	Treinamento			Teste		
	R2 adjust	RMSEu	RMSEs	R2 adjust	RMSEu	RMSEs
RLM	0.81	43.30	159.15	0.22	125.73	370.469
MLP	0,99	0.01	0.04	0.56	230.93	304.763
RF	0.49	191.65	78.08	0.30	455.21	164.532
XGB	0,99	0.01	0.02	0.54	267.53	284.775

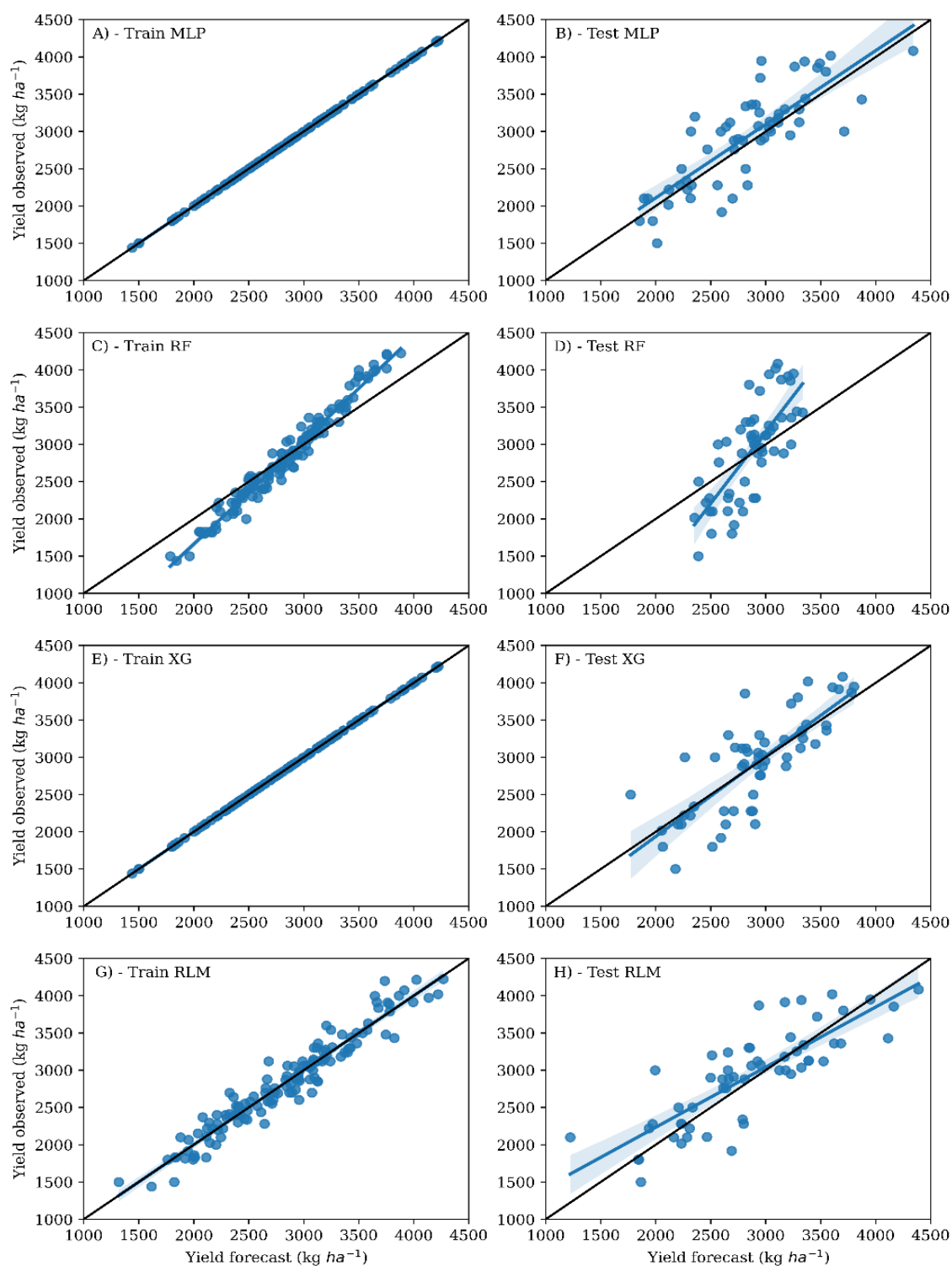


Figura 6. Desempenho entre os dados reais e previstos pelos algoritmos na calibração da previsão da soja

5. CONCLUSÃO

A produtividade da soja pode ser prevista de maneira precisa por meio da aplicação de algoritmos de aprendizado de máquina em dados climáticos. As variáveis de irradiância solar, temperatura média, máxima e mínima apresentou correlações positivas nos estádios finais de desenvolvimento da soja.

A regressão linear múltipla foi o algoritmo com menor acurácia e baixa precisão para predição da produtividade da soja utilizando o clima, apresentando valores de R^2_{adjust} , RMSEu e RMSEs de 0.22 kg ha⁻¹, 125.73 kg/ha e 370.469 kg/ha respectivamente. O modelo com maior acurácia, foi o Multi-layer Perceptron, tanto no período de treinamento e de teste com R^2_{adjust} , RMSEu e RMSEs de 0.99, 0.01 kg ha⁻¹, 0.04 kg ha⁻¹ e 0.56, 230.93 kg ha⁻¹ e 304.763 kg ha⁻¹ respectivamente. A MLP foi robusta, pois alcançou resultados precisos utilizando apenas variáveis climáticas. Isso também demonstra a importância dos fatores climáticos na produtividade da soja nas regiões da Bahia, Mato Grosso e São Paulo.

REFERÊNCIAS

- ABRAHAM, E. R. et al. Neural Network System to Forecast the Soybean Exportation on Brazilian Port of Santos. **IFIP Advances in Information and Communication Technology**, v. 514, 2017. 83-90 p.
- ABRAHÃO, G. M.; COSTA, M. H. Evolution of rain and photoperiod limitations on the soybean growing season in Brazil: The rise (and possible fall) of double-cropping systems. **Agricultural and Forest Meteorology**, v. 256-257, 2018. 32-45 p.
- AIKEN, L. S. et al. **Multiple linear regression**. [S.l.]: Handbook of Psychology, 2012.
- AKAMATSU, H. et al. Pathogenic diversity of soybean rust in Argentina, Brazil, and Paraguay. [S.l.]: **J Gen Plant Pathol**, 2013. 28–40 p.
- ALLEN, R. G. et al. Evapotranspiração das culturas — diretrizes para calcular as necessidades de água das culturas. **Documento da FAO sobre irrigação e drenagem 56**, Roma, 1998.
- ALLIPRANDINI, L. F. et al. Understanding Soybean Maturity Groups in Brazil: Environment, Cultivar Classification, and Stability. **Madison: Crop Science Society of America**, 2009.
- ALMEIDA, A. M. D. **Eficiência do uso da água e resposta termal da cultura da soja submetida a diferentes níveis de reposição hídrica**. Piracicaba: Universidade de São Paulo, 2022.
- APARECIDO, L. E. D. O. et al. Modelagem da produtividade do milho safrinha em função das condições climáticas do Mato Grosso do Sul. [S.l.]: **Revista Brasileira de Climatologia**, 2020.
- ARSHAD, S. et al. Exploring dynamic response of agrometeorological droughts towards winter wheat yield loss risk using machine learning approach at a regional scale in Pakistan. [S.l.]: **Field Crops Research**, 2023.
- ARVOR, D. et al. Análise de perfis temporais de EVI/MODIS para o monitoramento da cultura da soja no Estado do Mato Grosso, Brasil. Florianópolis : Anais **XIII Simpósio Brasileiro de Sensoriamento Remoto**, 2007.
- BARBOSA, D. A. et al. Influência do déficit hídrico sobre parâmetros agronômicos das cultivares de soja Embrapa 48 e BR 16 em condições de campo. [S.l.]: **Embrapa Soja**, 2015.
- BORGES, P. H. D. M. et al. Normal climatological water balance in the microregion of Alto Teles Pires - MT. [S.l.]: **Brazilian Applied Science Review**, 2023.

BORRMANN, D. **Efeito do déficit hídrico em características químicas e bioquímicas da soja e na manipulação da clorofila, com ênfase na formação de metabólitos incolores**. São Paulo: Faculdade de Ciências Farmacêuticas. 2009.

Cidade Brasil. Disponível em:< <https://www.cidade-brasil.com.br/>>. Acesso em 03 de Setembro de 2023.

CONAB. Acompanhamento da safra brasileira de grãos. **Décimo segundo levantamento - Safra 2022/2023**, Setembro 2023. 1-110 p.

DECARLI, L. et al. Seed quality and soybean plasticity with different growth habits. **Research, Society and Development**, v.11, 2022.

Embrapa Soja. **Soja em números - safra 2022/23**. Disponível em: <<https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>>. Acesso em 03 de Outubro de 2023.

ESCOBAR, O. D. S. et al. Severidade de *Phakopsora pachyrhizi* em soja estimada por modelos de crescimento não lineares. Passo Fundo: **Agrometeoros Revista Brasileira de Agrometeorologia**, v. 31, 2023.

FARIAS, J. R. B.; NEPOMUCENO, A. L.; NEUMAIER, N. **Ecofisiologia da Soja**. Londrina. 2007.

FEHR, W. R.; CAVINESS, C. E. **Stages of soybean development**. Ames, Iowa: State University of Science and Technology, 1977. 12 p.

FERREIRA, B. S.; GOMEZ, D. R. S. Percevejos e o sistema de produção soja-milho. Londrina: **Embrapa Soja**, 2017.

FERREIRA, J. C. S. **Grupos de maturação da cultura da soja avaliados por imagens aéreas obtidas por vant**. Rio Verde: Instituto Federal de Educação, Ciência e Tecnologia Goiano, 2020.

FERREIRA, J. F. K. **Comportamento do frete rodoviário de soja: aplicação de um modelo de regressão linear múltipla**. Campinas: [s.n.], 2016.

FILHO, D. F. et al. O que Fazer e o que Não Fazer com a Regressão: pressupostos e aplicações do modelo linear de Mínimos Quadrados Ordinários (MQO). [S.l.]: **Revista Política Hoje**, 2011.

FISHMAN, J. et al. An investigation of widespread ozone damage to the soybean crop in the upper Midwest determined from ground-based and satellite measurements. [S.l.]: **Atmospheric Environment**, 2010.

FOERSTER, A. M. L. **Períodos de interferência de plantas daninhas em culturas agrícolas**. ITACOATIARA: Instituto de Ciências Exatas e Tecnologia da Universidade Federal do Amazonas, 2021.

FRAJACOMO, H. C. **Seleção de SNPs utilizando Random Forests**, São Carlos, 2020.

FREITAS, I. L. D. J. **Fitotecnia: soja e milho**. Londrina: Editora e Distribuidora Educacional S.A, 2019.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5^a. ed. New York: AMGH Editora Ltda, 2011.

JUNIOR, W. D. C. et al. Multiple linear regression and Random Forest model to estimate soil bulk density in mountainous regions. [S.l.]: **Pesquisa Agropecuária Brasileira**, v. 51, 2016. 1428-1437 p.

KASTER, M.; FARIAS, J. R. B. Regionalização dos testes de valor de cultivo e uso e da indicação de cultivares de soja – terceira aproximação. Londrina: **Embrapa Soja**, 2012.

KNORR, M. T. **Quarenta anos de expansão da soja no Brasil, 1975-2015**. Disponível em: <<http://journals.openedition.org/confins/12592>>. Acesso em 15 de dezembro de 2022.

LANGENBACH, G. et al. Fighting Asian Soybean Rust. [S.l.]: **Frontiers in Plant Science**, v. 7, 2016.

LIU, X. et al. Geographic differentiation and phylogeographic relationships among world soybean populations. **The Crop Journal**, 8, n. 2, Abril 2020. 260-272.

MACHADO, H. G.; MUNDIM, K. Revisão das Tecnologias de Inteligência Artificial e Machine/Deep Learning: Restrições, Oportunidades, Estado da Arte e Desafios. [S.l.]: **Revista Processos Químicos**, 2023.

MANDARINO, J. M. G.; ROESSING, A. C. Tecnologia para produção do óleo de soja: descrição das etapas, equipamentos, produtos e subprodutos. Londrina: **Embrapa Soja**, 2001.

MARTELETTO, S. R. **Técnicas de seleção de atributos através de Random Forests: um estudo de caso para detecção de tendências em séries temporais financeiras**. São Paulo: Universidade de São Paulo, 2022.

MARTIN, T. N.; PIRES, J. L. F.; VEY, R. T. **Tecnologias Aplicadas para o Manejo Rentável e Eficiente da Cultura da Soja**. Santa Maria: Editora GR, 2022.

MENESES, K. C. D. Estimating Potential Evapotranspiration in Maranhão State Using Artificial Neural Networks. [S.l.]: **Revista Brasileira de Meteorologia**, 2020.

MIRANDA, J. M. et al. Modelo matemático para previsão da produtividade do cafeeiro. Campina Grande: **Revista Brasileira de Engenharia Agrícola e Ambiental**, 2013.

MONTEIRO, J. E. B. A. **Agrometeorologia dos cultivos: o fator meteorológico na produção agrícola**. Brasília: INMET, 2009.

MONTEIRO, L. A.; SENTELHAS, P. C. Calibration and testing of an agrometeorological model for the estimation of soybean yields in different Brazilian regions. **Acta Scientiarum - Agronomy**, 36, n. 3, 2014. 265-272.

NARAYANAKUMAR, S.; RAJA, K. A BP Artificial Neural Network Model for Earthquake Magnitude Prediction in Himalayas, India. [S.l.]: **Scientific Research**, 2016.

NEUMAIER, N. et al. Estádios de desenvolvimento da cultura de soja. Passo Fundo: **Embrapa Soja**, 2000. 44 p.

NOVAIS, G. T. Climas do Brasil: classificação climática e aplicações. Porto Alegre: **Totalbooks**, 2023.

OBSIE, E. Y.; QU, H.; DRUMMOND, F. Wild blueberry yield prediction using a combination of computer simulation and machine learning algorithms. [S.l.]: **Computers and Electronics in Agriculture**, 2020.

OLIVEIRA, M. F. et al. Analyzing complexities in the Brazilian soybean supply chain: a systems thinking and modeling approach. **RAUSP Management Journal**, 57, n. 3, 2022. 280-297 p.

OSMAN, A. I. A. et al. Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia. **Ain Shams Engineering Journal**, v. 12, n. 2, p. 1545-1556, 2021.

PEÑA-GALLARDO, M. et al. Response of crop yield to different time-scales of drought in the United States: Spatio-temporal patterns and climatic and environmental drivers. [S.l.]: **Agricultural and Forest Meteorology**, v. 264, 2019. 40-55 p.

PEREIRA, A. R.; ANGELOCCI, L. R.; SENTELHAS, P. C. Agrometeorologia: fundamentos e aplicações práticas. Guaíba: **Agropecuária**, 2002.

QUADROS, A. S. D. et al. Inoculação e coinoculação com *Bradyrhizobium* e *Azospirillum* na cultura da soja. [S.l.]: **Brazilian Journal of Animal and Environmental Research**, 2020.

QUIROZ, J. C. et al. "Fault detection of broken rotor bar in LS-PMSM using random forests. [S.l.]: **Measurement**, 2018.

RITCHIE, S. W. et al. Como a planta de soja se desenvolve. Piracicaba: **Potafós**, 1994.

RODRIGUES, O. et al. **Resposta quantitativa do florescimento da soja à temperatura e ao fotoperíodo**. Passo Fundo: Fisiologia Vegetal, 2000.

ROLIM, G. D. S. **Ajuste e comparação de modelos de simulação da produtividade para a determinação do risco climático da cultura do girassol (*Helianthus annuus* L.)**. Piracicaba: Universidade de São Paulo, 2000.

ROSA, V. D. R. et al. Estresse hídrico durante a fase reprodutiva de duas linhagens de soja. Brasília: **Pesquisa Agropecuária Brasileira**, 2020.

SANTOS, L. F. D. A. **Classificador baseado em Redes Neurais Convolucionais para algoritmos RMLSA em EONs**. Brasília: Universidade de Brasília, 2020.

SANTOS, V. B. D.; SANTOS, A. M. F. D.; ROLIM, G. D. S. Estimation and forecasting of soybean yield using artificial neural networks. **Agronomy Journal**, 113, n. 4, Julho 2021. 3193-3209 p.

SEIXAS, C. D. S. et al. Tecnologias de Produção de Soja. Londrina: **Embrapa Soja**, 2020.

SENTELHAS, P. C. et al. Clima e produtividade da soja: variabilidade climática como fator controlador da produtividade. Rondonópolis: **Boletim de Pesquisa de Soja**, 2017. 25-41 p.

SERAFIM, M. E. et al. Umidade do solo e doses de potássio na cultura da soja. Fortaleza: **Revista Ciência Agronômica**, v. 43, 2012. 222-227 p.

SILVA, A. G. D. et al. Predicting the quality of soybean seeds stored in different environments and packaging using machine learning. **Scientific Reports**, 12, n. 1, 2022.

SPANCERSKI, J. S.; SANTOS, J. A. A. D. Previsão da produtividade de arroz: uma aplicação de redes neurais recorrentes LSTM. [S.l.]: **Revista Cereus**, 2021.

TEJO, D. P.; FERNANDES, C. H. D. S.; BURRATO, J. S. Soja: fenologia, morfologia e fatores que interferem na produtividade. Garça: **Revista Científica Eletrônica XIX da FAEF**, 2019.

THOMAS, A. L.; COSTA, J. A. **Soja: manejo para alta produtividade de grãos**. Porto Alegre: Evangraf, 2010.

TILMAN, D. et al. **Global food demand and the sustainable intensification of agriculture**. [S.l.]: PNAS, 2011.

TRINDADE, F. S. **Monitoramento da cultura da soja em agrossistema com espectrorradiômetro orbital de resolução moderada**. Lavras: Universidade Federal de Lavras, 2018.

TRZECIAK, M. B. **Produtividade e qualidade fisiológica de sementes de soja na região de Pelotas-RS**. Pelotas: Universidade, 2009.

TSUKAHARA, R. Y. et al. Soybean yield as a consequence of harvest delay and environmental conditions. [S.l.]: **Pesquisa Agropecuária Brasileira**, 2016.

USDA. **Oilseeds and Products Update**. Brazil. 2023.

_____. **United States Agricultural Export Yearbook**. [S.l.]. 2023.

VALE, N. K. A. D.; ABDALA, K. D. O. Modelo teórico para a compreensão da interação entre as variáveis que interferem na produtividade da soja. Araguaia: **Revista Eletrônica de Educação da Faculdade Araguaia**, 2018.

XAVIER, A. et al. Using unsupervised learning techniques to assess interactions among complex traits in soybeans. [S.l.]: **Springer Science Business Media**, 2017.

YUZHASKANDI, S. S.; KHALILIAN, S. On Projecting Climate Change Impacts on Soybean Yield in Iran: an Econometric Approach. **Environmental Processes**, v.7, 2020. 73-87 p.